Yanan Liu[a,b,1], Na Li[c,1], Xiao Zhu[a,b,d,e,**], Yi Qi[a,b,d,e,*]

[a] Southern Marine Science and Engineering Guangdong Laboratory (Zhanjiang), Zhanjiang, China
[b] Guangdong Key Laboratory for Research and Development of Natural Drugs, The Marine Biomedical Research Institute, Guangdong Medical University, Zhanjiang, China
[c] College of Basic Medicine, Chongqing Medical University, Chongqing, China
[d] The Key Lab of Zhanjiang for R&D Marine Microbial Resources in the Beibu Gulf Rim, Guangdong Medical University, Zhanjiang, China
[e] The Marine Biomedical Research Institute of Guangdong Zhanjiang, Zhanjiang, China

ARTICLEINFO

Keywords:
Genetic big data
Precision medicine
Cancer
Neurological and psychiatric disorders
Angiocardiopathy
Artificial intelligence

ABSTRACT

In the era of big data, massive genetic data, as a new industry, has quickly swept almost all industries, especially the pharmaceutical industry. As countries around the world start to build their own gene banks, scientists study the data to explore the origins and migration of humans. Moreover, big data encourage the development of cancer therapy and bring good news to cancer patients. Big datum has been involved in the study of many diseases, and it has been found that analyzing diseases at the gene level can lead to more beneficial treatment options than ordinary treatments. This review will introduce the development of extensive data in medical research from the perspective of big data and tumor, neurological and psychiatric diseases, cardiovascular diseases, other applications and the development direction of big data in medicine.

1. Introduction

The application of genetic big data is widespread in biomedicine. With the development of various life technologies, the cost of exploring all the genomes in the human body has dropped sharply, and more genetic testing has served in numerous fields, including preventing diseases and exploring the pathogenesis of a transmissible disease. The stage of anthropological exploration gets out of individuals to groups, and from persons to DNA molecules. Genetic considerable data is the vast trend of the future development of life science.

With the continuous progress and development of gene big data, the study of the human genome has opened a new chapter at the molecular level. Because of its small population, Iceland became the first country to obtain its genes sequences. Some of the findings were published. The study revealed that certain genes influence normal physiological functions of the body [1]. In addition, scientists have found a link between mutated genes and disease through the Icelandic gene bank [2]. They also found rare mutations linked to Alzheimer disease (AD) [3]. By studying the entire gene of more than 700 men in Iceland, the mutation frequency of the male Y chromosome was studied. It utilizes to trace human origins [4]. A company called DECODE has sequence the genomes of more than 10,000 Icelanders and has tentatively mapped some 150,000 of them depended on their genetic relationships. In addition, a genome program in the UK has been undertaken a database of British people for treatment and research. The program has been applied to diagnose rare diseases and cancers [5]. Iceland and Britain have a strong genetic relationship with a small number of people. We can make preliminary inferences about genetic profiling and the treatment of human diseases based on the genomes of a few people.

Because Singapore is an important immigrant country, the race is abundant. Based on their nationality, the researchers can shed light on the evolutionary history of Asian populations [6]. They used the
Novo-Zhonghua database for mutation frequency annotation, which can avoid experimental errors caused by population differences and make the research results more accurate [7]. In China, the most authoritative genetic big data research results posted. It not only reveals that innate differences between the han nationality and other ethnic minorities but also Chinese inborn characteristics have been greatly influenced by history including the silk road and mass migration. In addition, different latitude and longitude have a profound influence on Chinese diet and daily life [8]. The latest article, published in *Nucleic Acids Research*, caused to stir. The researchers analyzed data on single nucleotide variations in Han genomes from 33 provinces, including some overseas Chinese. Not only that, the team also released a free online analysis of Han genetic database platform "PGG.Han" on its website, http://www. pgghan.org. They analyzed the genetic pool of 100,000 Han people and provided a free online analysis platform [9]. The Human Genome Project has not only identified many genes in the human body, but also accumulated a large amount of genetic data. With the help of big data, people can compare the development process and etiology of various diseases at the genetic level and study human

Abbreviations: AD, Alzheimer's disease; CNV, copy number variation; tRCC, translocation renal cell carcinoma; CCLE, Cancer Cell Line Encyclopedia; MSI, microsatellite instability; PD-L1, programmed death-ligand 1; TMB, tumor mutation burden; PD-1, programmed cell death protein 1; GWAS, genome-wide association study; HCV, hepatitis C virus; IFNγ, interferon γ.

* Corresponding author at: Guangdong Key Laboratory for Research and Development of Natural Drugs, The Marine Biomedical Research Institute, Guangdong Medical University, Zhanjiang, China; and Southern Marine Science and Engineering Guangdong Laboratory (Zhanjiang), Zhanjiang, China.
** Corresponding author at: Guangdong Key Laboratory for Research and Development of Natural Drugs, The Marine Biomedical Research Institute, Guangdong Medical University, Zhanjiang, China; and Southern Marine Science and Engineering Guangdong Laboratory (Zhanjiang), Zhanjiang, China. *E-mail addresses:* xzhu@gdmu.edu.cn (X. Zhu), qiyi7272@gdmu.edu.cn (Y. Qi).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.biopha.2020.111074
Received 22 September 2020; Received in revised form 16 November 2020; Accepted 27 November 2020
Available online 9 December 2020
0753-3322/© 2020 The Author(s). Published by Elsevier Masson SAS. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

growth and physiological activities at the molecular level. In this review, we introduce the research development of genetic big data about cancer, neurological and psychiatric diseases, to understand, prevent, and treat diseases better.

2. Genetic big data and tumors

2.1. Nature of cancer

2.1.1. Cancer generation

In 2015, Cristian Tomasetti and Bert Vogelstein of Johns Hopkins University found that most cancers are random [10,11]. A year later, by analyzing more than 4.8 billion samples from 69 countries around the world, they confirmed the role of random DNA replication errors in cancer: 28.9 percent of cancers were caused by environmental factors, 5 percent by genetic factors, and 66.1 percent by random DNA replication errors. This has upended what people always thought that about cancer.

Scientists at the time believed that tumors were caused primarily by random genetic errors in DNA replication [12].

After extensive analysis, scientists generally agree that the promoter of early cancer is a mutation activated by the RAS gene [13]. RAS, found in prostate cancer cells, was the first known oncogene [14]. By 2015, scientists had found more than 500 oncogenes [15]. In fact, the RAS mutation is the first step in a smoker becoming a lung cancer patient [16].

Depending on current research, if RAS mutates, it is the publisher of all commands. It activates signaling pathways and allows cells to secrete proteins, which promotes cell proliferation, growth and micro vascular formation, and leads to cell cancer. What's more, it causes cells to release special substances that attract macrophages, and eventually the cancer cells metastasize [17] (Fig. 1).

2.1.2. Mutation characteristics

Mutations are an issue in oncology because of the variety and complexity of the forms of mutations. Whole gene sequencing and integrated analysis based on different cancer types are important means for scientists to study cancer. Copy number variation (CNV) is one of the most common methods for clinical study of tumors. Depending on the CNV model, scientists can identify the type of cancer and the progression of the disease [18]. Here are some examples to illustrate. First, researchers combined various bioinformatics methods to analyze patients' exon sequences with translocation renal cell carcinoma (tRCC), and studied gene copy number changes and tumor mutation load. They found that copy number changes were prevalent in tRCC which may cause poor prognosis [19]. Second, the team proposed and created the first large-scale whole-genome sequencing of cancer. Studying 2658 cancers, the early and late stages of 38 cancers were re-identified. In addition, scientists reconstructed the process of mutation and the factors that caused it [20].

The current findings on what drives the mutations are incomplete and related men only can evaluate data based on fewer samples. The main purpose of this research is to promote the generation and development of PCAWG genomic resources through computing clouds and makes universal data sharing. The team assessed the entire genome of more than 2000 tumors from 38 tumor types, showing that the PCAWG project made considerable advance. Besides, the researchers discovered a little about driving mutations [21].

Recently, by analyzing the drive mutations and structural mutations in 2658 non-coding regions of the tumor genome, Rheinbay's team developed an algorithmic system to identify the mutations and found new drive mutations [22].

The study combined all the mutations which have been identified so far. While a lot of causes of numerous mutations are unknown, scientists can look for them by trying the effect of each mutation in a single cancer genome [23].

2.1.3. RNA change

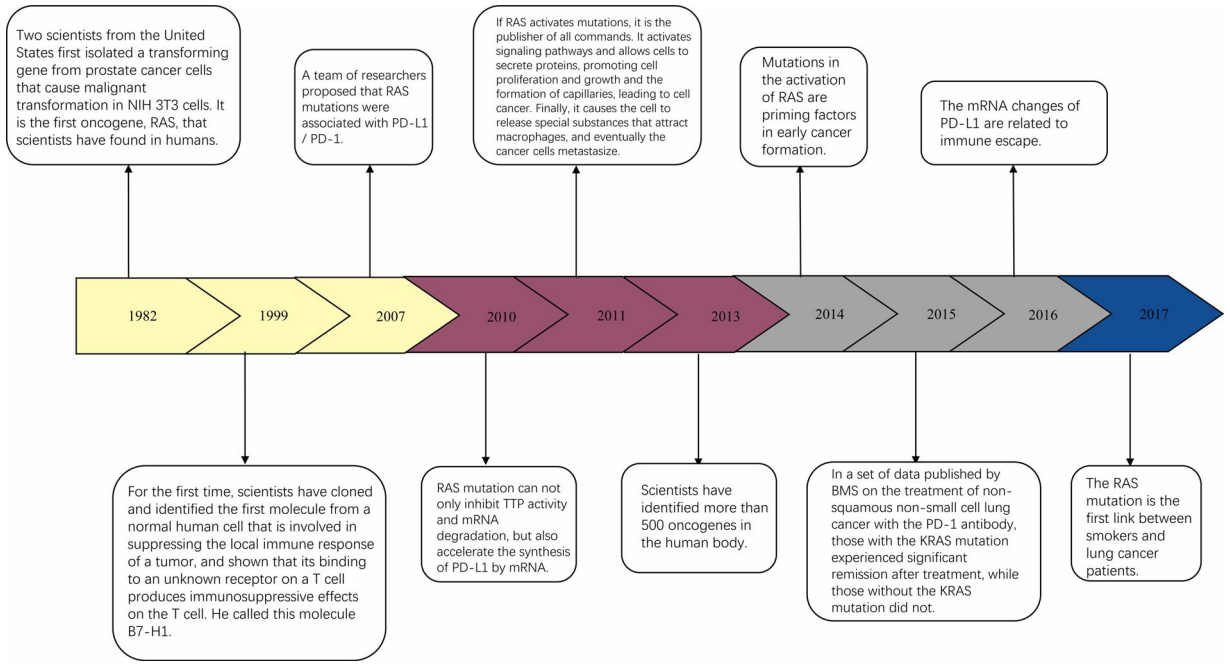
Two scientists from the United States first isolated a transforming gene from prostate cancer cells that cause malignant transformation in NIH 3T3 cells. It is the first oncogene, RAS, that scientists have found in humans.
A team of researchers proposed that RAS mutations were associated with PD-L1 / PD-1.
If RAS activates mutations, it is the publisher of all commands. It activates signaling pathways and allows cells to secrete proteins, promoting cell proliferation and growth and the formation of capillaries, leading to cell cancer. Finally, it causes the cell to release special substances that attract macrophages, and eventually the cancer cells metastasize.
Mutations in the activation of RAS are priming factors in early cancer formation.
The mRNA changes of PD-L1 are related to immune escape.
1982
1999
2007
2010
2011
2013
2014
2015
2016
2017
For the first time, scientists have cloned and identified the first molecule from a normal human cell that is involved in suppressing the local immune response of a tumor, and shown that its binding to an unknown receptor on a T cell produces immunosuppressive effects on the T cell. He called this molecule B7-H1.
RAS mutation can not only inhibit TTP activity and mRNA degradation, but also accelerate the synthesis of PD-L1 by mRNA.
Scientists have identified more than 500 oncogenes in the human body.
In a set of data published by BMS on the treatment of non-squamous non-small cell lung cancer with the PD-1 antibody, those with the KRAS mutation experienced significant remission after treatment, while those without the KRAS mutation did not.
The RAS mutation is the first link between smokers and lung cancer patients.

Fig. 1. Since its discovery, the RSA gene has been the most promising cancer gene in oncology and research. For 35 years, scientists have been studying cancer at the genetic level. They proved to be the strongest of all oncogenes. RAS activation mutation is linked with PD-1 / PD-L1. In cancer cells with RAS, RAS mutations inhibit TTP activity and mRNA degradation. In addition, RAS mutation can also accelerate mRNA synthesis of PD-L1. It ends up connecting cancer cells to a lot of PD-L1. In future cancer studies, RAS mutations are liable to be used as markers to predict the efficacy of pD-1 / PD-L1 antibody drugs. This is of major help to patients treated with checkpoint inhibitors.

More and more people are studying tumors at the genetic level. By linking RNA changes to DNA mutations, the team suggests possible genetic mechanisms. Because different RNA has different effects on different mutations, the biology is quite different. In fact, different RNA changes can provide researchers with directions to identify genes and mechanisms associated with cancer [24].

2.1.4. Structural variation

The key mutation process in cancer is structural variation, which can range from thousands of basic groups to entire chromosomes. The study, which analyzed the entire genome of more than 2600 tumors, identified structural changes in 16 types of cancer. Because genes are rearranged when they are placed, complex genome configurations are generated [25,26].

2.1.5. Anticancer drugs

Researchers searched public databases for tumor samples. By comparing the numerous cancer cells in the lab, they looked for cancer lines with genetic mutations associated with the samples, which are equivalent to mimic how a drug acts on cancer cells. According to the principal characteristics of cancer cell lines in the clinic, it is possible to find drugs to fight this cancer [27–31].

2.2. Tumors example

Lu daopei hospital is devoted to the diagnosis of leukemia and has found the distribution and basic characteristics of fusion genes. Furthermore, differences in Leukemia among patients of different races and ages are analyzed. This study provides a mechanism for the occurrence and development of leukemia from the molecular perspective, and makes a significant contribution to improving the clinical fusion gene detection program [32]. Now researchers have developed a precise leukemia vaccine. Through the self-healing properties of the microcapsules, the vaccine can increase the level of specific cytotoxic T lymphocyte activation [33]. With further study of leukemia, scientists have found that the drug Arsenic Trioxide is able to treat medulloblastoma [34].

For breast cancers, the researchers identified 191 risk genes [35] in the 150 regions of human DNA [36] that are known to be susceptible to breast cancer. Some researchers have found that breast cancer patients must use dietary supplements with caution, or they will increase the risk of recurrence. They analyzed more than 1000 patients with breast cancer and gave them different drug tests before and during chemotherapy [37].

Single positive breast cancer is a special type of breast cancer. A medical team from China assessed the characteristics of cancer through big data analysis. The endocrine effect of estrogen receptor single positive is worse than that of estrogen progesterone receptor double positive breast cancer. Furthermore, 75 % of estrogen receptor-positive breast cancer patients had copy number loss or promoter methylation of the progesterone receptor gene, which together explained the progesterone receptor-negative effect [38].

The most common tumor suppressor gene in head and neck cancer is TP53, which can influence tumor formation by regulating tumor microenvironment. TP53 is active in the process of nerve regeneration to control the nerves at the stage of differentiation [39]. By comparing cancer-related neurons in a mouse model of oral cancer, the researchers found that nervous system drugs were effective in treating head and neck tumors [40]. The study revealed that head and neck cancer was influenced by the human papillomavirus and smoking. In addition, the researchers compared head and neck cancer with other cancers, which could help other scientists in the upcoming genetic age [41].

In non-invasive prenatal testing, the cisco discovery protocol can be utilized to diagnose occult maternal malignancy in pregnant women with multiple chromosome aneuploidies (better when combined with plasma tumor markers). Researchers diagnosed a pregnant woman with small-cell cancer seven years ago. Her fetus was polyploid on chromosomes 13 and 18, but the karyotype was normal. Five years ago, researchers also performed genetic tests on more than 120, 000 pregnant women, and 10 of them were identified as having malignant tumors during pregnancy [42–44]. *2.3. Encyclopedia of cancer cell lines*

Siu LL et al. used big data to analyze cancer cells and build cancer networks with other countries and medical institutions [45]. The team has extracted the cancer susceptibility genome from the tumor genome [46].

While previous studies of cancer have focused on just 1% of the protein code, this study takes the lead in the expression of 99 % of the genes. Researchers from 37 countries around the world studied different stages and properties of cancer and analyzed more than 2800 genomes to build a database of cancer genes for humans. In the paper, the researchers analyzed nearly all the biological pathways and their role in cancer. These can help clinicians detect early cancers and accurately identify cancer types [21,47–49]. Tumors with the same gene drive can result in different outcomes in different patients because the cancer drives different genes [26,50–52].

The researchers performed an integrated analysis of 2658 whole cancer genomes and a pan-cancer analysis of the entire cancer genome of normal tissue to 38 tumor types that they matched. By aggregating tumor sequences with the human genome and analyzing mutations in human somatic cells, they identified previously unknown genetic mutations associated with cancer types [21].

Cancer Cell Line Encyclopedia (CCLE) is a large data set that includes gene expression, DNA copy Numbers, and other data from nearly 1000 Cell lines of different lineages [53]. The researchers analyzed the cell lines in CCLE by mass spectrometry and found a correlation between protein expression and biological pathways.

Then, they used big data to study the relationship between microsatellite instability (MSI) in colorectal cancer and proteome and found the effect on transcription [54]. Protein histology is still in its development stage, and the highest level is the quantitative analysis of most expressed proteins in hundreds of samples.

Although tumor progression varies from species to species, scientists can use the classification of

data mining to build models to predict cancers. Taking silicon lung cancer as an example, scientists have studied the development of cancer by analyzing different models, such as animal experiments [55]. *2.4. Tumor detection and treatment*

In cancer detection, biomarkers, such as programmed death-ligand 1 (PD-L1), MSI, Tumor mutation burden (TMB) and so on, accurately identify patients who respond to drugs through rapid and direct detection. They are all the contemporary mainstream biomarkers used to detect cancer. By comparing the overall response rate after drug administration, while ensuring extremely high sensitivity and positive predictive rate, PD-L1 has very good specificity and positive predictive rate, which are suitable for cancer detection in all aspects [56].

PD-L1 was the first molecule to become involved in human's local immunosuppression of tumors, when it was called B7-H1 [57]. In a study, scientists used antibody programmed cell death protein 1 (PD-1) on the treatment of non-squamous non-small cell lung cancer, and subjects with the KRAS gene mutation experienced significant remission after treatment [58]. It has been found that as long as more than half of the cancer cells express PD-L1, many indications acquire antibodies PD-1 [59–62]. In fact, the progressive rate of PD-L1 in patients with KRAS gene is higher than that in patients without KRAS gene. Through big data analysis, there were some unknown relationships between RAS and PD-L1 [63,64]. Kataoka K's team from Kyoto university in Japan found that changes in PD-L1 mRNA were involved in immune escape [65].

RAS gene mutation keeps the stable signal of the mRNA encoding PD- L1 protein to the downstream signal, and then continuously gives the cancer cell to synthesize PD-L1 to resist PD-1 on the surface of T cells. It can be seen from this that the cancer cell can immunosuppress the attack of the immune system in the human body [66] (Fig. 1).

The accumulation of errors in the replication of genetic material by organisms is recognized as MSI. However, DNA mutations in tumor cells are mostly manifested in abnormal microsatellite sequences, which the immune cells immediately recognize as a problem. TMB is the number of mutations per megabyte of Tumor tissue. The principle is very close to MSI. Different cancers have different TMB levels, and MSI and TMB may show an impact on cross-cancer types in treatment [67].

In order to improve the efficiency of PD-L1/PD-1, scientists and clinicians discovered the king combination of Lenvatini and PD-1. Taking liver cancer as an example, the research team recruited 104 patients with liver cancer to receive the therapy of the combination. Researchers showed that the objective remission rate was significantly reduced. The survival cycle was significantly improved and the annual survival rate of the patients was 67.5 % [68]. In renal cancer, PD-1 resistance achieved a 51 percent effective rates and a 91 percent disease control rates in the study overall. In addition to the above data, it has likewise published data on non-small cell lung cancer, head and neck squamous cell carcinoma, melanoma [69].

3. Genetic big data with neurological and psychiatric disorders

The pathogenesis of autism is a lot more complex than we thought. The largest study of the Autism genome to date, led by Autism Speaks, tells us that most related people have different autism genes, according to research [70]. The study could prove that the human body is under a large number of genes that influence autism (Table 1).

For years, AD has been a therapeutic problem. Not only that, with the help of genetic big data, some scientists release their findings to the world. For one thing, they found a gene that can delay the onset of AD – a low expression of PU.1 protein [71]. Second, scientists from Iceland discovered a rare mutation associated with AD [3]. In three studies, researchers at Stanford university found immune T cells in the cerebrospinal fluid of patients with AD. In fact, the immune system plays a sizable role in the face of AD, especially in the neuroinflammatory process of adaptive immunity [72]. (Table 1).

The team identified 202 risk genes in more than 1.33 million British patients with insomnia and discovered 15 genes most likely to be involved in insomnia in the genome wide association study (GWAS). In addition, they identified the genes of 450,000 people and found 57 loci linked to insomnia. The findings are a starting point to treat insomnia, as well as other conditions [73]. In addition, Lane JM et al. found that frequent insomnia can affect aging, metabolism, reproduction and other shared genetic factors. They found that insomnia can cause symptoms such as depression and subjective well-being [74] (Table 1).

The study found that there was a 50 percent chance that diverticula

Table 1
The disease that concerns disease and nerve commonly is perceptual, is difficult to use medicaments to cure completely in medical respect. The drug that comes out at present is great majority is with restraint and alleviate give priority to. The discovery of genes could open the door to more cures.

Disease	Research data	Research results	URL	References
autism	publicity	Most related people with autism spectrum disorder (ASD) have different autism genes	https://www.ncbi.nlm.nih.gov/pubmed/25,621,899	RK et al. (2015), B et al. (2015)
		The researchers showed low expression of a gene called PU.1 that slows the onset of Alzheimer's disease There are rare mutations associated with Alzheimer's disease.	https://www.ncbi.nlm.nih.gov/pubmed/28,628,103 https://www.ncbi.nlm.nih.gov/pubmed/25,807,283	KL et al. (2017), E et al. (2017) S et al. (2015), H et al. (2015)
Alzheimer disease	private	Immune T cells have been found in cerebrospinal fluid in patients with Alzheimer's disease (AD). In fact, the immune system	https://www.ncbi.nlm.nih.gov/pubmed/31,915,375	D et al. (2020), N et al. (2020)

		plays a big role in the face of Alzheimer's disease.		
insomnia	private	The human body has 202 risk genes for insomnia, 57 loci associated with insomnia, and the 15 most likely to be associated with insomnia were speculated: DAB1, MEIS1, BTBD9, SDK1, TRPC7, TMEM161B, LEMD2, FOXP2, PTPRD, CNTNS, OLFM4, SNX29, RBFOX1, DDC, NOL4L	https://www.ncbi.nlm.nih.gov/pubmed/30,804,565	PR et al. (2019), K et al. (2019)
		Frequent insomnia can affect the aging, metabolism, reproduction and other shared genetic factors, and can make patients have depression, subjective well-being and other symptoms.	https://www.ncbi.nlm.nih.gov/pubmed/30,804,566	IM et al. (2019), SE et al. (2019)
diverticulosis	private	Genes associated with the action of some calcium channel blockers (non-dihydropyridines) in humans have a 50 % chance of being associated with diverticular disease.	https://www.ncbi.nlm.nih.gov/pubmed/31,234,639	D *et al.* (2019), MK et al. (2019)

disease was associated with genes involved in the action of some calcium channel blockers (non-dihydropyridines) [75]. If patients with high blood pressure take some calcium channel blockers for a long time, they may have intestinal problems in the body (Table 1).

4. Genetic big data and cardiovascular diseases

There are further genes in the body that affect cardiovascular disease. The researchers systematically analyzed the RNA sequence data of the blood, blood vessels, and metabolic tissues from patients with coronary artery disease, and determined the genes that partly caused the disease: some genes were expressed in a single tissue, and some genes were expressed across multiple tissues to cause cardiovascular disease.

Now, scientists are using these findings to redefine disease, its properties, and so on [76]. Italian residents who eat more chili pepper may be under a lower risk of cardiovascular disease. Studies have shown a negative correlation between eating chilies regularly and the risk of cardiovascular disease [77]. **5. Other applications of genetic big data**

5.1. Using genetic big data to define health and disease

Havrilla JM is part of a team that targeted the genes of more than 100,000 healthy people to identify regions that are not easily altered [78]. The study found that DNA mutations in "restricted" areas of the human body may result in serious childhood diseases.

The researchers tested the health of more than 100 participants at risk for the disease. Depending on the health curve, they can find some disease detection criteria and markers. They can capture the definition of health and the boundaries of disease [79]. Although the research is expensive, it will be cut as more people study it. Further research is continuing, and it is believed that in the future people will be in a position to go from treating illnesses to prevent them to keep everyone healthy. Everyone gets a large number of genetic mutations in their bodies. Some teams combine exon data with electronic medical records to analyze diseases at the gene level. They first analyzed the relationship between fasting lipid levels and rare mutations. Through a big data analysis, they found a relationship between variants of five genes, APOC3, APOB and lipid levels. If a person carries a fatal genetic mutation, he may be at risk of death [80,81]. In addition, because of obtaining more population genetic data, Khera AV et al. improved polygenic risk scoring. They plan to use this method to help high risk patients prevent and treat disease [82].

5.2. Advances in genetics, immunity and physiology of big data

The study of the Y chromosome could help biological historians explore hominid presence in Iceland. DECODE explored the frequency of mutations in the male Y chromosome by studying the entire genome of 700 men in Iceland. This mutation frequency can be utilized to trace human origins [4]. The paternal map of Chinese people was constructed and improved by means of the gene data on the Y chromosome of 2000 Chinese men. The researchers focused on the Y haploid group o2a1c-002,611 and inferred the distribution, origin and migration direction of Han and Tibetan populations in China at this Y haploid group [83].

Dutch scientists studied genetic data from more than 500 healthy Caucasians in Europe to explore the influence of non-genetic factors such as age, sex and environment of ontogeny. It focuses on changes in some cytokines produced by the immune system during aging, such as interleukin-22 (IL-22) and interferon γ (IFNγ). The concentration of IFNγ decreased in most people. This study demonstrates that immune function declines to a certain extent as people age [84]. In addition, other scientists working to build cell maps have found that genetic variation has a strong effect on gene expression in cells [28,51,85–90].

The subjects are about 70,000 cells in the maternal-fetal junction at six to 14 weeks of gestation, including placental trophoblastic cells and maternal deciduous cells. The team studied which genes were turned on by different cells according to their cell types and sequencing methods. Not only that, by identifying which genes are used during pregnancy, researchers analyze the role of certain genes. What's more, they can also understand the

whole process of pregnancy on a molecular level [91].

Although all organisms share a common genetic code, there are millions of different loci in individual's genetic code. Recently, a team from the Institute of immunology (gene) released a lot of data. The data could help researchers in the industry decipher the role of genetic variation in the immune system and detail how mutations in DNA sequences affect genetic activity in immune cells. They analyzed more than 1500 samples, screened the data and found that genetic activity in the same type of immune cells varied significantly by sex and age. In addition, they found that genetic variations affect the expression of nearby genes and so on. They used the data to build a map of human immune cells, proving that genetic variation has a strong effect on gene expression [85].

Mice have long been frequent subjects of gut research. The team from China has completed the mouse gut intestinal microbial resource bank through the large-scale isolation, culture and sequencing of the cecum bacteria in mice, and obtained the draft genomes of more than 100 species mice [92]. Papers published by Greenblum S and others made an open a sensation in the world. By studying and analyzing the microbiota and microbiota genes in the gut of mice, they eventually found genetic mutations in the microbiota [93]. Researchers who analyzed the composition of bacteria and lipids in detail found a direct link between the bacteria and obesity and diabetes. In fact, studies have shown that microbiota disorders may be associated with noncommunicable diseases such as cardiovascular disease in humans [94].

Genetic big data are widely used in clinical practice. The following is examples of recent advances in cell preparation using genetic big data *in vitro*.

A team from China has been working on a way to prepare mature human cells *in vitro*, with success so far. The team simulated the environment that developed *in vivo*, using a new way of lineage reprogramming to produce mature human cells *in vitro* [95].

The researchers created blood cells *in vitro* by studying the six stages of an early embryo. The entire network of gene interactions that control the development of blood cells has also been discovered [96].

5.3. Genetic big data in non-human species

The team, led by Ansari MA, set out to analyze viral hepatitis. They found a breakthrough in hepatitis C virus (HCV), how HCV interacts with its host [97]. Understanding the genetic basis that influences the outcome of infectious diseases will enable future treatments and prevention approaches to be targeted [86,87,98,99].

Microorganisms are almost impossible to grow in culture medium [100]. However, with the development of sequencing technology, a website called KOMODO combines the microbial classification and database to predict the formulation of culture medium based on the sequence of 16S rDNA of microorganisms. Scientists gradually combined the microbial data based on 16S rDNA with the culture of the strain, which promoted the microbial research and resource saving [101].

6. The development direction of genetic big data

6.1. AI and medicine

Currently, scientists are exploring the interaction between medicine and AI, hoping to continuously improve the health system and the doctor-patient relationship in the future [102,103]. Recently, paper on cloud computing has caused a lot of discussion. The team analyzed facial features of more than 17,000 patients with genetic disorders and used deep learning to identify the link between facial features and genetic disorders.

Big data analysis helps users to sort and summarize complex data according to certain needs, especially in the medical field. The analysis methods of big data are basically applicable to the management of public health system and the analysis of medical data, such as deep learning and cloud computing [104].

Graph embedding has been applied to biomedical network analysis [105]. The 11 kinds of graph of embedding techniques are summarized in 3 categories, which are MF based method, random walk based method and the neural network based method. Even more impressive, the authors' research team used data mining from open clinical texts to categorize the semantic types of medical terms and to model useful information [106]. Not only that, network topology structure is also well applied in biomedical field [105].

Now they have developed an algorithm called DeepGestalt that can efficiently identify rare genetic diseases [107]. In fact, a model of diabetes-related kidney disease that uses real-world data is a better predictor than one that uses clinical data in the laboratory [108]. Anyone doing research can't just look at clinical data and ignore real-world data.

Big data are just a simple term for ordinary people. Big breakthroughs are possible if big data are combined with problems that are currently hard for people to solve [109]. Deep learning is a hot technology in the study of medical image analysis. In clinical analysis, the direction of obtaining medical images by learning network has changed from partial dependence on deep learning to registration based on supervised learning. Different deep learning frameworks improve the accuracy of registration task [110]. For interpretable AI to be oriented toward primary care, the algorithm must have visibility [111]. Machine learning is a hot research topic these days. By analyzing the data, computers can build appropriate models to help people make predictions. But the process is not necessarily precise. Interactive machine learning is about connecting experts with modeling. Expertise and big data analysis provides more accurate predictions for users [112]. (Fig. 2)

However, medical big data itself are deceptive. In medical research, there are academic frauds, such as falsifying data and tampering with data. Modeling vulnerabilities also pose risks for data warehouses [113]. Sometimes the interference of centralizing data can cause model output error [114].

6.2. Gene therapeutics

Scientists have discovered the possibility of using gene therapy to cure Gaucher disease. Using a mouse model, the team inserted normal genes into the genetic code of the virus and then injected them into mice. Remarkably, the researchers found the mutation and fixed it. The research also offers treatments for genetic diseases related to metabolism [30,115].

6.3. Single-cell sequencing

The development of hematopoietic and immune systems in mammals, including humans, during embryonic development is influenced by many tissues and organs [116–118]. It is because of ethical and moral factors that prevent people from studying human organs. Many scientists use mice for research because their embryonic organs have highly similar development patterns to those of humans [119]. Popescu DM et al. successfully mapped the development of human embryonic hematopoietic and immune system in single cell sequencing [120].

6.4. DNA storage

Because of its special structure, DNA has the potential to store a lot of information. Scientists have come up with a storage strategy called DNA fountain, which uses DNA to store computer operating systems. 215PB of data are stored in each gram of DNA and retrieved by sequencing [121].

7. Conclusions and challenges

Genetic big data are to analyze and study the deep meaning of the diseases we want to explore through the analysis of a large number of samples and the combination and continuity between samples. Because
everyone's genes are different, even if they have the same disease, the clinical symptoms are different from person to person. Doctors must devise different treatment plans according to the situation. This is known as "personalized medicine". Experts divide a disease into many subtypes and classify treatment, which is called "precision medicine" [26,103,122,123].

The researchers studied more than 56 m mutated genes in more than 340,000 unrelated people in the UK. They screened for nine of these traits and found that the environment enhanced or reduced the effect of genotypes on body mass index (BMI) and other obesity-related genes [124]. Not only that, habits can also be influenced by genes. Using genetic big data, the researchers found a group of genes in the body's DNA that influence the ability to get up early, and found that people who get up early have better mental states than those who stay up late [125]. They analyzed whole-genome and circadian rhythm data from more than 700,000 people and found 351 genomes that regulate the body clock by participating in the cAMP signaling pathway, glutamate and insulin signaling pathways. In addition, these genomes are mainly distributed in the retina, pituitary gland and hypothalamus.

Big data are actually used as a means of clinical research to help scientists study the pathogenesis of diseases and other harm to human body. It is a new technology in the new era. Scientists use genetic analysis to study genes in the body that act like drugs to explore their effects. Of course, these findings are not intended to change clinical practice, but to spur further research.

Funding

This work was supported partly by Guangdong Science and Technology Department (2016B030309002 and 2019B090905011); The Fund of Southern Marine Science and Engineering Guangdong Laboratory (Zhanjiang) (ZJW-2019-007); and the Science and Technology Program of Zhanjiang (2017A06012); The Grant of Chongqing Education Commission (KJQN202000406); The Public Service Platform of South China Sea for R&D Marine Biomedicine Resources (GDMUK201808).

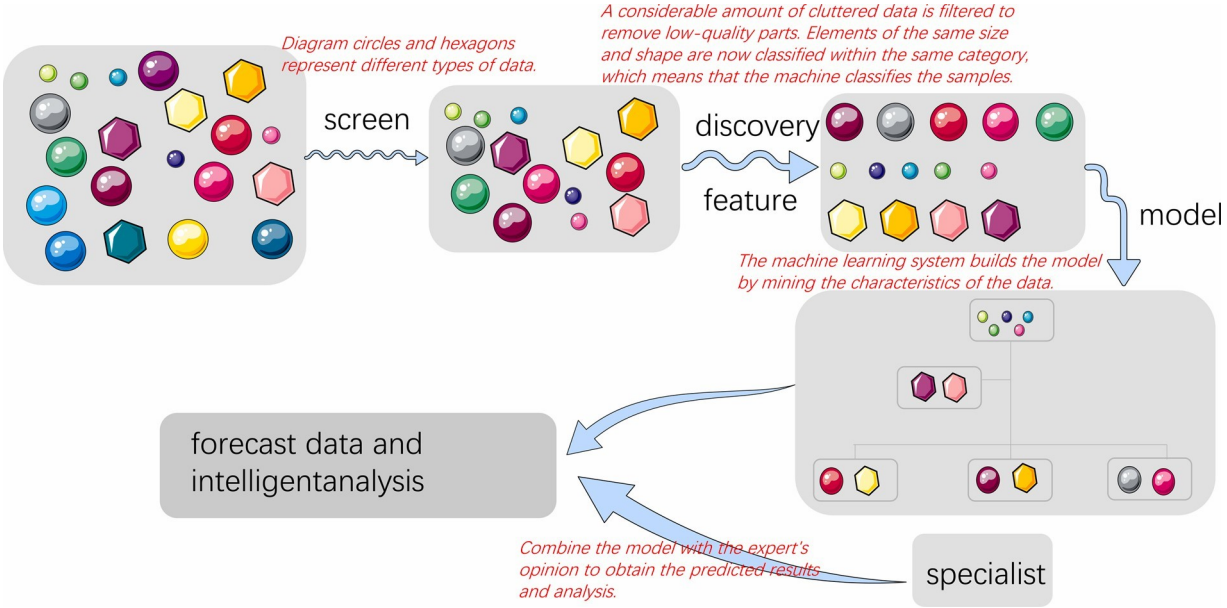
Diagram circles and hexagons represent different types of data.
screen
A considerable amount of cluttered data is filtered to remove low-quality parts. Elements of the same size and shape are now classified within the same category, which means that the machine classifies the samples.
discovery
feature
model
The machine learning system builds the model by mining the characteristics of the data.
forecast data and intelligentanalysis
Combine the model with the expert's opinion to obtain the predicted results and analysis.
specialist

Fig. 2. The application process of deep learning in medicine. A data model comprises of a large number of data volumes. They are intricate and interconnected. Researchers gather data from a variety of sources (researchers should pay attention to data privacy issues) and further standardize and integrate the data as needed. Then the machine learning system can recognize and weigh the characteristics of the data, excavate the characteristic patterns from the data, and finally build the deep learning model. Because the existing deep learning model is not very mature, the output is only the predicted results and lack of analytical process. At present, the establishment of deep learning model still needs the participation of experts to design the optimal structure.

Authors' contributions

XZ conceived the work. YL wrote and drafted the manuscript. NL and XZ discussed and edited the manuscript. YQ and XZ funded the study. All authors read and approved the final version of the manuscript.

Declaration of Competing Interest

The authors report no declarations of interest.

References

[1] P. Sulem, H. Helgason, A. Oddson, H. Stefansson, S.A. Gudjonsson, F. Zink, E. Hjartarson, G.T. Sigurdsson, A. Jonasdottir, A. Jonasdottir, A. Sigurdsson, O. T. Magnusson, A. Kong, A. Helgason, H. Holm, U. Thorsteinsdottir, G. Masson, D. F. Gudbjartsson, K. Stefansson, Identification of a large set of rare complete human knockouts, Nat. Genet. 47 (5) (2015) 448–452.

[2] D.F. Gudbjartsson, H. Helgason, S.A. Gudjonsson, F. Zink, A. Oddson, A. Gylfason, S. Besenbacher, G. Magnusson, B.V. Halldorsson, E. Hjartarson, G. T. Sigurdsson, S.N. Stacey, M.L. Frigge, H. Holm, J. Saemundsdottir, H. T. Helgadottir, H. Johannsdottir, G. Sigfusson, G. Thorgeirsson, J.T. Sverrisson, S. Gretarsdottir, G.B. Walters, T. Rafnar, B. Thjodleifsson, E.S. Bjornsson, S. Olafsson, H. Thorarinsdottir, T. Steingrimsdottir, T.S. Gudmundsdottir, A. Theodors, J.G. Jonasson, A. Sigurdsson, G. Bjornsdottir, J.J. Jonsson, O. Thorarensen, P. Ludvigsson, H. Gudbjartsson, G.I. Eyjolfsson, O. Sigurdardottir, I. Olafsson, D.O. Arnar, O.T. Magnusson, A. Kong, G. Masson, U. Thorsteinsdottir, A. Helgason, P. Sulem, K. Stefansson, Large-scale whole- genome sequencing of the Icelandic population, Nat. Genet. 47 (5) (2015) 435–444.

[3] S. Steinberg, H. Stefansson, T. Jonsson, H. Johannsdottir, A. Ingason, H. Helgason, P. Sulem, O.T. Magnusson, S.A. Gudjonsson, U. Unnsteinsdottir, A. Kong, S. Helisalmi, H. Soininen, J.J. Lah, D.Aarsland DemGene, T. Fladby, I. D. Ulstein, S. Djurovic, S.B. Sando, L.R. White, G.P. Knudsen, L.T. Westlye, G. Selbaek, I. Giegling, H. Hampel, M. Hiltunen, A.I. Levey, O.A. Andreassen, D. Rujescu, P.V. Jonsson, S. Bjornsson, J. Snaedal, K. Stefansson, Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease, Nat. Genet. 47 (5) (2015) 445–447.

[4] A. Helgason, A.W. Einarsson, V.B. Guethmundsdottir, A. Sigurethsson, E. D. Gunnarsdottir, A. Jagadeesan, S.S. Ebenesersdottir, A. Kong, K. Stefansson, The Y-chromosome point mutation rate in humans, Nat. Genet. 47 (5) (2015) 453–457.

[5] M. Peplow, The 100,000 genomes project, BMJ 353 (2016) i1757.

[6] M. Cohen, A. Giladi, A.D. Gorki, D.G. Solodkin, M. Zada, A. Hladik, A. Miklosi, T. M. Salame, K.B. Halpern, E. David, S. Itzkovitz, T. Harkany, S. Knapp, I. Amit, Lung single-cell signaling interaction map reveals basophil role in macrophage imprinting, Cell 175 (4) (2018) 1031–1044, e18.

[7] W. Zheng, H. Chen, X. Deng, L. Yuan, Y. Yang, Z. Song, Z. Yang, Y. Wu, H. Deng, Identification of a novel mutation in the Titin Gene in a chinese family with limb- girdle muscular dystrophy 2J, Mol. Neurobiol. 53 (8) (2016) 5097–5102.

[8] S. Liu, S. Huang, F. Chen, L. Zhao, Y. Yuan, S.S. Francis, L. Fang, Z. Li, L. Lin, R. Liu, Y. Zhang, H. Xu, S. Li, Y. Zhou, R.W. Davies, Q. Liu, R.G. Walters, K. Lin, J. Ju, T. Korneliussen, M.A. Yang, Q. Fu, J. Wang, L. Zhou, A. Krogh, H. Zhang, W. Wang, Z. Chen, Z. Cai, Y. Yin, H. Yang, M. Mao, J. Shendure, J. Wang, A. Albrechtsen, X. Jin, R. Nielsen, X. Xu, Genomic analyses from non-invasive prenatal testing reveal genetic associations, patterns of viral infections, and chinese population history, Cell 175 (2) (2018) 347–359, e14.

[9] Y. Gao, C. Zhang, L. Yuan, Y. Ling, X. Wang, C. Liu, Y. Pan, X. Zhang, X. Ma, Y. Wang, Y. Lu, K. Yuan, W. Ye, J. Qian, H. Chang, R. Cao, X. Yang, L. Ma, Y. Ju, L. Dai, Y. Tang, K.I. Han, G. Zhang, S. Xu, PGG.Han: the Han Chinese genome database and analysis platform, Nucleic Acids Res. 48 (D1) (2020) D971–D976.

[10] C. Tomasetti, B. Vogelstein, Cancer risk: role of environment-response, Science 347 (6223) (2015) 729–731.

[11] C. Tomasetti, B. Vogelstein, Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions, Science 347 (6217) (2015) 78–81.

[12] C. Tomasetti, L. Li, B. Vogelstein, Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention, Science 355 (6331) (2017) 1330–1334.

[13] A.D. Cox, S.W. Fesik, A.C. Kimmelman, J. Luo, C.J. Der, Drugging the undruggable RAS: Mission possible? Nat. Rev. Drug Discov. 13 (11) (2014) 828–851.

[14] M. Goldfarb, K. Shimizu, M. Perucho, M. Wigler, Isolation and preliminary characterization of a human transforming gene from T24 bladder carcinoma cells, Nature 296 (5856) (1982) 404–409.

[15] B. Vogelstein, N. Papadopoulos, V.E. Velculescu, S. Zhou, L.A. Diaz Jr., K. W. Kinzler, Cancer genome landscapes, Science 339 (6127) (2013) 1546–1558.

[16] M. Vaz, S.Y. Hwang, I. Kagiampakis, J. Phallen, A. Patil, H.M. O'Hagan, L. Murphy, C.A. Zahnow, E. Gabrielson, V.E. Velculescu, H.P. Easwaran, S. B. Baylin, Chronic cigarette smoke-induced epigenomic changes precede sensitization of bronchial epithelial cells to single-step transformation by KRAS mutations, Cancer Cell 32 (3) (2017) 360–376, e6.

[17] Y. Pylayeva-Gupta, E. Grabocka, D. Bar-Sagi, RAS oncogenes: weaving a tumorigenic web, Nat. Rev. Cancer 11 (11) (2011) 761–774.

[18] T. Huang, Copy number variations in tumors, Ref. Module Biomed. Sci. (2018) 10–12.

[19] J. Marcon, R.G. DiNatale, A. Sanchez, R.R. Kotecha, S. Gupta, F. Kuo, V. Makarov, A. Sandhu, R. Mano, A.W. Silagy, K.A. Blum, D.E. Nassau, N.E. Benfante, M. V. Ortiz, M.I. Carlo, T.A. Chan, R.J. Motzer, M.H. Voss, J. Coleman, P. Russo, V. Reuter, A.A. Hakimi, E. Reznik, Comprehensive genomic analysis of translocation renal cell carcinoma reveals copy-number variations as drivers of disease progression, Clin. Cancer Res. 26 (14) (2020) 3629–3640.

[20] M. Gerstung, C. Jolly, I. Leshchiner, S.C. Dentro, S. Gonzalez, D. Rosebrock, T. J. Mitchell, Y. Rubanova, P. Anur, K. Yu, M. Tarabichi, A. Deshwar, J. Wintersinger, K. Kleinheinz, I. Vazquez-Garcia, K. Haase, L. Jerman, S. Sengupta, G. Macintyre, S. Malikic, N. Donmez, D.G. Livitz, M. Cmero, J. Demeulemeester, S. Schumacher, Y. Fan, X. Yao, J. Lee, M. Schlesner, P. C. Boutros, D.D. Bowtell, H. Zhu, G. Getz, M. Imielinski, R. Beroukhim, S. C. Sahinalp, Y. Ji, M. Peifer, F. Markowetz, V. Mustonen, K. Yuan, W. Wang, Q. D. Morris, P. Evolution, G. Heterogeneity Working, P.T. Spellman, D.C. Wedge, P. Van Loo, P. Consortium, The evolutionary history of 2,658 cancers, Nature 578 (7793) (2020) 122–128.

[21] Consortium I.T.P.-C.Ao.W.G, Pan-cancer analysis of whole genomes, Nature 578 (7793) (2020) 82–93.

[22] E. Rheinbay, M.M. Nielsen, F. Abascal, J.A. Wala, O. Shapira, G. Tiao, H. Hornshoj, J.M. Hess, R.I. Juul, Z. Lin, L. Feuerbach, R. Sabarinathan, T. Madsen, J. Kim, L. Mularoni, S. Shuai, A. Lanzos, C. Herrmann, Y.E. Maruvka, C. Shen, S.B. Amin, P. Bandopadhayay, J. Bertl, K.A. Boroevich, J. Busanovich, J. Carlevaro-Fita, D. Chakravarty, C.W.Y. Chan, D. Craft, P. Dhingra, K. Diamanti, N.A. Fonseca, A. Gonzalez-Perez, Q. Guo, M.P. Hamilton, N.J. Haradhvala,

C. Hong, K. Isaev, T.A. Johnson, M. Juul, A. Kahles, A. Kahraman, Y. Kim,
J. Komorowski, K. Kumar, S. Kumar, D. Lee, K.V. Lehmann, Y. Li, E.M. Liu,
L. Lochovsky, K. Park, O. Pich, N.D. Roberts, G. Saksena, S.E. Schumacher,
N. Sidiropoulos, L. Sieverling, N. Sinnott-Armstrong, C. Stewart, D. Tamborero, J.
M.C. Tubio, H.M. Umer, L. Uuskula-Reimand, C. Wadelius, L. Wadi, X. Yao, C.
Z. Zhang, J. Zhang, J.E. Haber, A. Hobolth, M. Imielinski, M. Kellis, M.
S. Lawrence, C. von Mering, H. Nakagawa, B.J. Raphael, M.A. Rubin, C. Sander, L. D. Stein, J.M. Stuart, T. Tsunoda, D.A. Wheeler, R. Johnson, J. Reimand,
M. Gerstein, E. Khurana, P.J. Campbell, N. Lopez-Bigas, P. Drivers, G. Functional
Interpretation Working, P.S.V.W. Group, J. Weischenfeldt, R. Beroukhim,
I. Martincorena, J.S. Pedersen, G. Getz, P. Consortium, Analyses of non-coding somatic drivers in 2,658 cancer whole genomes, Nature 578 (7793) (2020) 102–111.

[23] L.B. Alexandrov, J. Kim, N.J. Haradhvala, M.N. Huang, A.W. Tian Ng, Y. Wu, A. Boot, K.R. Covington, D.A. Gordenin, E.N. Bergstrom, S.M.A. Islam, N. Lopez- Bigas, L.J. Klimczak, J.R. McPherson, S. Morganella, R. Sabarinathan, D.
A. Wheeler, V. Mustonen, P.M.S.W. Group, G. Getz, S.G. Rozen, M.R. Stratton, P. Consortium, The repertoire of mutational signatures in human cancer, Nature 578 (7793) (2020) 94–101.

[24] P.T.C. Group, C. Calabrese, N.R. Davidson, D. Demircioglu, N.A. Fonseca, Y. He,
A. Kahles, K.V. Lehmann, F. Liu, Y. Shiraishi, C.M. Soulette, L. Urban, L. Greger,
S. Li, D. Liu, M.D. Perry, Q. Xiang, F. Zhang, J. Zhang, P. Bailey, S. Erkek, K.
A. Hoadley, Y. Hou, M.R. Huska, H. Kilpinen, J.O. Korbel, M.G. Marin,
J. Markowski, T. Nandi, Q. Pan-Hammarstrom, C.S. Pedamallu, R. Siebert, S.
G. Stark, H. Su, P. Tan, S.M. Waszak, C. Yung, S. Zhu, P. Awadalla, C.J. Creighton,
M. Meyerson, B.F.F. Ouellette, K. Wu, H. Yang, P.T.W. Group, A. Brazma, A.
N. Brooks, J. Goke, G. Ratsch, R.F. Schwarz, O. Stegle, Z. Zhang, P. Consortium, Genomic basis for RNA alterations in cancer, Nature 578 (7793) (2020) 129–136.

[25] Y. Li, N.D. Roberts, J.A. Wala, O. Shapira, S.E. Schumacher, K. Kumar, E. Khurana, S. Waszak, J.O. Korbel, J.E. Haber, M. Imielinski, P.S.V.W. Group, J. Weischenfeldt, R. Beroukhim, P.J. Campbell, P. Consortium, Patterns of somatic structural variation in human cancer genomes, Nature 578 (7793) (2020) 112–121.

[26] Y. Li, T. Tao, L. Du, X. Zhu, Three-dimensional genome: developmental technologies and applications in precision medicine, J. Hum. Genet. 65 (6) (2020) 497–511.

[27] F. Iorio, T.A. Knijnenburg, D.J. Vis, G.R. Bignell, M.P. Menden, M. Schubert,
N. Aben, E. Goncalves, S. Barthorpe, H. Lightfoot, T. Cokelaer, P. Greninger,
E. van Dyk, H. Chang, H. de Silva, H. Heyn, X. Deng, R.K. Egan, Q. Liu,
T. Mironenko, X. Mitropoulos, L. Richardson, J. Wang, T. Zhang, S. Moran,
S. Sayols, M. Soleimani, D. Tamborero, N. Lopez-Bigas, P. Ross-Macdonald,
M. Esteller, N.S. Gray, D.A. Haber, M.R. Stratton, C.H. Benes, L.F.A. Wessels, J. Saez-Rodriguez, U. McDermott, M.J. Garnett, A landscape of pharmacogenomic interactions in cancer, Cell 166 (3) (2016) 740–754.

[28] K. Li, H. Luo, L. Huang, H. Luo, X. Zhu, Microsatellite instability: a review of what the oncologist should know, Cancer Cell Int. 20 (2020) 16.

[29] Z. Wei, L. Huang, L. Cui, X. Zhu, Mannose: Good player and assister in pharmacotherapy, Biomed. Pharmacother. 129 (2020), 110420.

[30] X. Liang, D. Li, S. Leng, X. Zhu, RNA-based pharmacotherapy for tumors: from bench to clinic and back, Biomed. Pharmacother. 125 (2020), 109997.

[31] S. Li, Z. Zhang, W.-F. Lai, L. Cui, X. Zhu, How to overcome the side effects of tumor immunotherapy, Biomed. Pharmacother. 130 (2020), 110639.

[32] X. Chen, F. Wang, Y. Zhang, M. Wang, W. Tian, W. Teng, X. Ma, L. Guo, J. Fang, Y. Zhang, P. Zhu, H. Liu, Panoramic view of common fusion genes in a large cohort of Chinese de novo acute myeloid leukemia patients, Leuk. Lymphoma 60 (4) (2019) 1071–1078.

[33] X. Xie, Y. Hu, T. Ye, Y. Chen, L. Zhou, F. Li, X. Xi, S. Wang, Y. He, X. Gao, W. Wei, G. Ma, Y. Li, Therapeutic vaccination against leukaemia via the sustained release of co-encapsulated anti-PD-1 and a leukaemia-associated antigen, Nat. Biomed. Eng. (2020), https://doi.org/10.1038/s41551-020-00624-6. Online ahead of print.

[34] P.H. Dos Santos Klinger, L.E.A. Delsin, G.A.V. Cruzeiro, A.F. Andrade, R.C.P. Lira, P.V. de Andrade, P.F. das Chagas, R.G. de Paula Queiroz, F.A. Trevisan, R.S. de Oliveira, C.A. Scrideli, L.G. Tone, E.T. Valera, Arsenic Trioxide exerts cytotoxic and radiosensitizing effects in pediatric Medulloblastoma cell lines of SHH Subgroup, Sci. Rep. 10 (1) (2020) 6836.

[35] L. Fachal, H. Aschard, J. Beesley, D.R. Barnes, J. Allen, S. Kar, K.A. Pooley,
J. Dennis, K. Michailidou, C. Turman, P. Soucy, A. Lemacon, M. Lush, J.P. Tyrer,
M. Ghoussaini, M. Moradi Marjaneh, X. Jiang, S. Agata, K. Aittomaki, M.
R. Alonso, I.L. Andrulis, H. Anton-Culver, N.N. Antonenkova, A. Arason, V. Arndt,
K.J. Aronson, B.K. Arun, B. Auber, P.L. Auer, J. Azzollini, J. Balmana, R.
B. Barkardottir, D. Barrowdale, A. Beeghly-Fadiel, J. Benitez, M. Bermisheva,
K. Bialkowska, A.M. Blanco, C. Blomqvist, W. Blot, N.V. Bogdanova, S.E. Bojesen,
M.K. Bolla, B. Bonanni, A. Borg, K. Bosse, H. Brauch, H. Brenner, I. Briceno, I.
W. Brock, A. Brooks-Wilson, T. Bruning, B. Burwinkel, S.S. Buys, Q. Cai, T. Caldes,
M.A. Caligo, N.J. Camp, I. Campbell, F. Canzian, J.S. Carroll, B.D. Carter, J.
E. Castelao, J. Chiquette, H. Christiansen, W.K. Chung, K.B.M. Claes, C.L. Clarke,
G.S. Collaborators, E. Collaborators, J.M. Collee, S. Cornelissen, F.J. Couch,
A. Cox, S.S. Cross, C. Cybulski, K. Czene, M.B. Daly, M. de la Hoya, P. Devilee,
O. Diez, Y.C. Ding, G.S. Dite, S.M. Domchek, T. Dork, I. Dos-Santos-Silva, A. Droit,
S. Dubois, M. Dumont, M. Duran, L. Durcan, M. Dwek, D.M. Eccles, C. Engel,
M. Eriksson, D.G. Evans, P.A. Fasching, O. Fletcher, G. Floris, H. Flyger,
L. Foretova, W.D. Foulkes, E. Friedman, L. Fritschi, D. Frost, M. Gabrielson,
M. Gago-Dominguez, G. Gambino, P.A. Ganz, S.M. Gapstur, J. Garber, J.
A. Garcia-Saenz, M.M. Gaudet, V. Georgoulias, G.G. Giles, G. Glendon, A.
K. Godwin, M.S. Goldberg, D.E. Goldgar, A. Gonzalez-Neira, M.G. Tibiletti, M.
H. Greene, M. Grip, J. Gronwald, A. Grundy, P. Guenel, E. Hahnen, C.A. Haiman,
N. Hakansson, P. Hall, U. Hamann, P.A. Harrington, J.M. Hartikainen,
M. Hartman, W. He, C.S. Healey, B.A.M. Heemskerk-Gerritsen, J. Heyworth,
P. Hillemanns, F.B.L. Hogervorst, A. Hollestelle, M.J. Hooning, J.L. Hopper,
A. Howell, G. Huang, P.J. Hulick, E.N. Imyanitov, K.C. Investigators,
H. Investigators, A. Investigators, C. Isaacs, M. Iwasaki, A. Jager, M. Jakimovska,
A. Jakubowska, P.A. James, R. Janavicius, R.C. Jankowitz, E.M. John, N. Johnson, M.E. Jones, A. Jukkola-Vuorinen, A. Jung, R. Kaaks, D. Kang, P.
M. Kapoor, B.Y. Karlan, R. Keeman, M.J. Kerin, E. Khusnutdinova, J.I. Kiiski,
J. Kirk, C.M. Kitahara, Y.D. Ko, I. Konstantopoulou, V.M. Kosma, S. Koutros,
K. Kubelka-Sabit, A. Kwong, K. Kyriacou, Y. Laitman, D. Lambrechts, E. Lee,
G. Leslie, J. Lester, F. Lesueur, A. Lindblom, W.Y. Lo, J. Long, A. Lophatananon, J.
T. Loud, J. Lubinski, R.J. MacInnis, T. Maishman, E. Makalic, A. Mannermaa,

M. Manoochehri, S. Manoukian, S. Margolin, M.E. Martinez, K. Matsuo,
T. Maurer, D. Mavroudis, R. Mayes, L. McGuffog, C. McLean, N. Mebirouk,
A. Meindl, A. Miller, N. Miller, M. Montagna, F. Moreno, K. Muir, A.M. Mulligan,
V.M. Munoz-Garzon, T.A. Muranen, S.A. Narod, R. Nassir, K.L. Nathanson, S.
L. Neuhausen, H. Nevanlinna, P. Neven, F.C. Nielsen, L. Nikitina-Zake,
A. Norman, K. Offit, E. Olah, O.I. Olopade, H. Olsson, N. Orr, A. Osorio, V.
S. Pankratz, J. Papp, S.K. Park, T.W. Park-Simon, M.T. Parsons, J. Paul, I.
S. Pedersen, B. Peissel, B. Peshkin, P. Peterlongo, J. Peto, D. Plaseska-Karanfilska,
K. Prajzendanc, R. Prentice, N. Presneau, D. Prokofyeva, M.A. Pujana, K. Pylkas, P. Radice, S.J. Ramus, J. Rantala, R. Rau-Murthy, G. Rennert, H.A. Risch,
M. Robson, A. Romero, M. Rossing, E. Saloustros, E. Sanchez-Herrero, D.
P. Sandler, M. Santamarina, C. Saunders, E.J. Sawyer, M.T. Scheuner, D.
F. Schmidt, R.K. Schmutzler, A. Schneeweiss, M.J. Schoemaker, B. Schottker,
P. Schurmann, C. Scott, R.J. Scott, L. Senter, C.M. Seynaeve, M. Shah, P. Sharma,
C.Y. Shen, X.O. Shu, C.F. Singer, T.P. Slavin, S. Smichkoska, M.C. Southey, J.
J. Spinelli, A.B. Spurdle, J. Stone, D. Stoppa-Lyonnet, C. Sutter, A.J. Swerdlow, R.
M. Tamimi, Y.Y. Tan, W.J. Tapper, J.A. Taylor, M.R. Teixeira, M. Tengstrom, S.
H. Teo, M.B. Terry, A. Teule, M. Thomassen, D.L. Thull, M. Tischkowitz, A.
E. Toland, R. Tollenaar, I. Tomlinson, D. Torres, G. Torres-Mejia, M.A. Troester,
T. Truong, N. Tung, M. Tzardi, H.U. Ulmer, C.M. Vachon, C.J. van Asperen, L.
E. van der Kolk, E.J. van Rensburg, A. Vega, A. Viel, J. Vijai, M.J. Vogel, Q. Wang, B. Wappenschmidt, C.R. Weinberg, J.N. Weitzel, C. Wendt, H. Wildiers,
R. Winqvist, A. Wolk, A.H. Wu, D. Yannoukakos, Y. Zhang, W. Zheng, D. Hunter,
P.D.P. Pharoah, J. Chang-Claude, M. Garcia-Closas, M.K. Schmidt, R.L. Milne, V. N. Kristensen, J.D. French, S.L. Edwards, A.C. Antoniou, G. Chenevix-Trench, J. Simard, D.F. Easton, P. Kraft, A.M. Dunning, Fine-mapping of 150 breast cancer risk regions identifies 191 likely target genes, Nat. Genet. 52 (1) (2020) 56–73. [36] K. Michailidou, S. Lindstrom, J. Dennis, J. Beesley, S. Hui, S. Kar, A. Lemacon,
P. Soucy, D. Glubb, A. Rostamianfar, M.K. Bolla, Q. Wang, J. Tyrer, E. Dicks,
A. Lee, Z. Wang, J. Allen, R. Keeman, U. Eilber, J.D. French, X. Qing Chen,
L. Fachal, K. McCue, A.E. McCart Reed, M. Ghoussaini, J.S. Carroll, X. Jiang,
H. Finucane, M. Adams, M.A. Adank, H. Ahsan, K. Aittomaki, H. Anton-Culver, N. N. Antonenkova, V. Arndt, K.J. Aronson, B. Arun, P.L. Auer, F. Bacot,
M. Barrdahl, C. Baynes, M.W. Beckmann, S. Behrens, J. Benitez, M. Bermisheva,
L. Bernstein, C. Blomqvist, N.V. Bogdanova, S.E. Bojesen, B. Bonanni, A.
L. Borresen-Dale, J.S. Brand, H. Brauch, P. Brennan, H. Brenner, L. Brinton,
P. Broberg, I.W. Brock, A. Broeks, A. Brooks-Wilson, S.Y. Brucker, T. Bruning,
B. Burwinkel, K. Butterbach, Q. Cai, H. Cai, T. Caldes, F. Canzian, A. Carracedo, B.
D. Carter, J.E. Castelao, T.L. Chan, T.Y. David Cheng, K. Seng Chia, J.Y. Choi, H. Christiansen, C.L. Clarke, N. Collaborators, M. Collee, D.M. Conroy,
E. Cordina-Duverger, S. Cornelissen, D.G. Cox, A. Cox, S.S. Cross, J.
M. Cunningham, K. Czene, M.B. Daly, P. Devilee, K.F. Doheny, T. Dork, I. Dos- Santos-Silva, M. Dumont, L. Durcan, M. Dwek, D.M. Eccles, A.B. Ekici, A.
H. Eliassen, C. Ellberg, M. Elvira, C. Engel, M. Eriksson, P.A. Fasching,
J. Figueroa, D. Flesch-Janys, O. Fletcher, H. Flyger, L. Fritschi, V. Gaborieau, M. Gabrielson, M. Gago-Dominguez, Y.T. Gao, S.M. Gapstur, J.A. Garcia-Saenz,
M.M. Gaudet, V. Georgoulias, G.G. Giles, G. Glendon, M.S. Goldberg, D.
E. Goldgar, A. Gonzalez-Neira, G.I. Grenaker Alnaes, M. Grip, J. Gronwald,
A. Grundy, P. Guenel, L. Haeberle, E. Hahnen, C.A. Haiman, N. Hakansson,
U. Hamann, N. Hamel, S. Hankinson, P. Harrington, S.N. Hart, J.M. Hartikainen,
M. Hartman, A. Hein, J. Heyworth, B. Hicks, P. Hillemanns, D.N. Ho,
A. Hollestelle, M.J. Hooning, R.N. Hoover, J.L. Hopper, M.F. Hou, C.N. Hsiung,
G. Huang, K. Humphreys, J. Ishiguro, H. Ito, M. Iwasaki, H. Iwata,
A. Jakubowska, W. Janni, E.M. John, N. Johnson, K. Jones, M. Jones, A. Jukkola- Vuorinen, R. Kaaks, M. Kabisch, K. Kaczmarek, D. Kang, Y. Kasuga, M.J. Kerin, S. Khan, E. Khusnutdinova, J.I. Kiiski, S.W. Kim, J.A. Knight, V.M. Kosma, V.
N. Kristensen, U. Kruger, A. Kwong, D. Lambrechts, L. Le Marchand, E. Lee, M.
H. Lee, J.W. Lee, C. Neng Lee, F. Lejbkowicz, J. Li, J. Lilyquist, A. Lindblom,
J. Lissowska, W.Y. Lo, S. Loibl, J. Long, A. Lophatananon, J. Lubinski,
C. Luccarini, M.P. Lux, E.S.K. Ma, R.J. MacInnis, T. Maishman, E. Makalic, K. E. Malone, I.M. Kostovska, A. Mannermaa, S. Manoukian, J.E. Manson,
S. Margolin, S. Mariapun, M.E. Martinez, K. Matsuo, D. Mavroudis, J. McKay, C. McLean, H. Meijers-Heijboer, A. Meindl, P. Menendez, U. Menon, J. Meyer,
H. Miao, N. Miller, N.A.M. Taib, K. Muir, A.M. Mulligan, C. Mulot, S.
L. Neuhausen, H. Nevanlinna, P. Neven, S.F. Nielsen, D.Y. Noh, B.
G. Nordestgaard, A. Norman, O.I. Olopade, J.E. Olson, H. Olsson, C. Olswold,
N. Orr, V.S. Pankratz, S.K. Park, T.W. Park-Simon, R. Lloyd, J.I.A. Perez,
P. Peterlongo, J. Peto, K.A. Phillips, M. Pinchev, D. Plaseska-Karanfilska,
R. Prentice, N. Presneau, D. Prokofyeva, E. Pugh, K. Pylkas, B. Rack, P. Radice,
N. Rahman, G. Rennert, H.S. Rennert, V. Rhenius, A. Romero, J. Romm, K.
J. Ruddy, T. Rudiger, A. Rudolph, M. Ruebner, E.J.T. Rutgers, E. Saloustros, D.
P. Sandler, S. Sangrajrang, E.J. Sawyer, D.F. Schmidt, R.K. Schmutzler, A. Schneeweiss, M.J. Schoemaker, F. Schumacher, P. Schurmann, R.J. Scott,
C. Scott, S. Seal, C. Seynaeve, M. Shah, P. Sharma, C.Y. Shen, G. Sheng, M.
E. Sherman, M.J. Shrubsole, X.O. Shu, A. Smeets, C. Sohn, M.C. Southey, J.
J. Spinelli, C. Stegmaier, S. Stewart-Brown, J. Stone, D.O. Stram, H. Surowy,
A. Swerdlow, R. Tamimi, J.A. Taylor, M. Tengstrom, S.H. Teo, M. Beth Terry, D.
C. Tessier, S. Thanasitthichai, K. Thone, R. Tollenaar, I. Tomlinson, L. Tong,
D. Torres, T. Truong, C.C. Tseng, S. Tsugane, H.U. Ulmer, G. Ursin, M. Untch, C. Vachon, C.J. van Asperen, D. Van Den Berg, A.M.W. van den Ouweland, L. van der Kolk, R.B. van der Luijt, D. Vincent, J. Vollenweider, Q. Waisfisz, S. Wang- Gohrke, C.R. Weinberg, C. Wendt, A.S. Whittemore, H. Wildiers, W. Willett,
R. Winqvist, A. Wolk, A.H. Wu, L. Xia, T. Yamaji, X.R. Yang, C. Har Yip, K.Y. Yoo,
J.C. Yu, W. Zheng, Y. Zheng, B. Zhu, A. Ziogas, E. Ziv, A. Investigators, A.
I. ConFab, S.R. Lakhani, A.C. Antoniou, A. Droit, I.L. Andrulis, C.I. Amos, F.
J. Couch, P.D.P. Pharoah, J. Chang-Claude, P. Hall, D.J. Hunter, R.L. Milne,
M. Garcia-Closas, M.K. Schmidt, S.J. Chanock, A.M. Dunning, S.L. Edwards, G. D. Bader, G. Chenevix-Trench, J. Simard, P. Kraft, D.F. Easton, Association analysis identifies 65 new breast cancer risk loci, Nature 551 (7678) (2017) 92–94.
[37] C.B. Ambrosone, G.R. Zirpoli, A.D. Hutson, W.E. McCann, S.E. McCann, W.
E. Barlow, K.M. Kelly, R. Cannioto, L.E. Sucheston-Campbell, D.L. Hershman, J.
M. Unger, H.C.F. Moore, J.A. Stewart, C. Isaacs, T.J. Hobday, M. Salim, G.
N. Hortobagyi, J.R. Gralow, G.T. Budd, K.S. Albain, Dietary supplement use during chemotherapy and survival outcomes of patients with breast cancer Enrolled in a

cooperative group clinical trial (SWOG S0221), J. Clin. Oncol. (2019) JCO1901203.
[38] X.Y. Liu, D. Ma, X.E. Xu, X. Jin, K.D. Yu, Y.Z. Jiang, Z.M. Shao, Genomic landscape and endocrine-resistant subgroup in estrogen receptor-positive, progesterone receptor-negative, and HER2-Negative breast cancer, Theranostics 8 (22) (2018) 6386–6399.
[39] M.H. Yun, P.B. Gates, J.P. Brockes, Regulation of p53 is critical for vertebrate limb regeneration, Proc. Natl. Acad. Sci. U. S. A. 110 (43) (2013) 17392–17397. [40] M. Amit, H. Takahashi, M.P. Dragomir, A. Lindemann, F.O. Gleber-Netto, C. R. Pickering, S. Anfossi, A.A. Osman, Y. Cai, R. Wang, E. Knutsen, M. Shimizu,
C. Ivan, X. Rao, J. Wang, D.A. Silverman, S. Tam, M. Zhao,
C. Caulin, A. Zinger, E. Tasciotti, P.M. Dougherty, A. El-Naggar, G.A. Calin, J.N. Myers, Loss of p53 drives neuron reprogramming in head and neck cancer, Nature 578 (7795)
(2019) 449–454.
[41] N. Cancer Genome Atlas, Comprehensive genomic characterization of head and neck squamous cell carcinomas, Nature 517 (7536) (2015) 576–582.
[42] X. Ji, J. Li, Y. Huang, P.L. Sung, Y. Yuan, Q. Liu, Y. Chen, J. Ju, Y. Zhou, S. Huang,
F. Chen, Y. Han, W. Yuan, C. Fan, Q. Zhao, H. Wu, S. Feng, W. Liu, Z. Li, J. Chen,
M. Chen, H. Yao, L. Zeng, T. Ma, S. Fan, J. Zhang, K.Y. Yuen, S.H. Cheng, I.W.
S. Chik, N.T. Liu, J. Zhu, S. Lin, J. Cao, S. Tong, Z. Shan, W. Li, M.R. Hekmat,
M. Garshasbi, J. Suela, Y. Torres, J.C. Cigudosa, F.J.P. Ruiz, L. Rodriguez,
M. Garcia, J. Bernik, E. Traven, U. Res, N. Tul, C.F. Tseng, D. Zhao, L. Sun, Q. Pan,
L. Shen, M. Dai, Y. Wang, J. Wang, H. Yang, Y. Yin, T. Duan, B. Zhu, M. Choolani, X. Jin, Y. Chen, M. Mao, Identifying occult maternal malignancies from 1.93 million pregnant women undergoing noninvasive prenatal screening tests, Genet. Med. 21 (10) (2019) 2293–2302.
[43] N.A. Pavlidis, Coexistence of pregnancy and malignancy, Oncologist 7 (4) (2002)
279–287.
[44] C.M. Osborne, E. Hardisty, P. Devers, K. Kaiser-Rogers, M.A. Hayden, W. Goodnight, N.L. Vora, Discordant noninvasive prenatal testing results in a patient subsequently diagnosed with metastatic disease, Prenat. Diagn. 33 (6) (2013) 609–611.
[45] L.L. Siu, M. Lawler, D. Haussler, B.M. Knoppers, J. Lewin, D.J. Vis, R.G. Liao, F. Andre, I. Banks, J.C. Barrett, C. Caldas, A.A. Camargo, R.C. Fitzgerald, M. Mao,
J.E. Mattison, W. Pao, W.R. Sellers, P. Sullivan, B.T. Teh, R.L. Ward, J.
C. ZenKlusen, C.L. Sawyers, E.E. Voest, Facilitating a culture of responsible and effective sharing of cancer genome data, Nat. Med. 22 (5) (2016) 464–471.
[46] S. Park, F. Supek, B. Lehner, Systematic discovery of germline cancer predisposition genes through the identification of somatic second hits, Nat. Commun. 9 (1) (2018) 2601.
[47] Y. Zhou, Y. Kong, W. Fan, T. Tao, Q. Xiao, N. Li, X. Zhu, Principles of RNA methylation and their implications for biology and medicine, Biomed.
Pharmacother. 131 (2020), 110731.
[48] B. Guo, D. Li, L. Du, X. Zhu, piRNAs: biogenesis and their potential roles in cancer, Cancer Metastasis Rev. 39 (2) (2020) 567–575.
[49] J. Liu, D. Li, H. Luo, X. Zhu, Circular RNAs: the star molecules in cancer, Mol.
Aspects Med. 70 (2019) 141–152.
[50] E. Porta-Pardo, L. Garcia-Alonso, T. Hrabe, J. Dopazo, A. Godzik, A pan-cancer catalogue of cancer driver protein interaction interfaces, PLoS Comput. Biol. 11 (10) (2015), e1004518.
[51] X. Zhu, H. Luo, Y. Xu, Transcriptome analysis reveals an important candidate gene involved in both nodal metastasis and prognosis in lung adenocarcinoma, Cell Biosci. 9 (2019) 92.
[52] G. Lu, H. Luo, X. Zhu, Targeting the GRP78 pathway for cancer therapy, Front.
Med. (Lausanne) 7 (2020) 351.
[53] J. Barretina, G. Caponigro, N. Stransky, K. Venkatesan, A.A. Margolin, S. Kim, C. J. Wilson, J. Lehar, G.V. Kryukov, D. Sonkin, A. Reddy, M. Liu, L. Murray, M.
F. Berger, J.E. Monahan, P. Morais, J. Meltzer, A. Korejwa, J. Jane-Valbuena, F. A. Mapa, J. Thibault, E. Bric-Furlong, P. Raman, A. Shipway, I.H. Engels,
J. Cheng, G.K. Yu, J. Yu, P. Aspesi Jr., M. de Silva, K. Jagtap, M.D. Jones, L. Wang, C. Hatton, E. Palescandolo, S. Gupta, S. Mahan, C. Sougnez, R.C. Onofrio,
T. Liefeld, L. MacConaill, W. Winckler, M. Reich, N. Li, J.P. Mesirov, S.B. Gabriel,
G. Getz, K. Ardlie, V. Chan, V.E. Myer, B.L. Weber, J. Porter, M. Warmuth,
P. Finan, J.L. Harris, M. Meyerson, T.R. Golub, M.P.
Morrissey, W.R. Sellers, R. Schlegel, L.A. Garraway, The cancer cell line encyclopedia enables predictive modelling of anticancer drug sensitivity, Nature 483 (7391) (2012) 603–607.
[54] D.P. Nusinow, J. Szpyt, M. Ghandi, C.M. Rose, E.R. McDonald 3rd, M. Kalocsay, J. Jane-Valbuena, E. Gelfand, D.K. Schweppe, M. Jedrychowski, J. Golji, D.
A. Porter, T. Rejtar, Y.K. Wang, G.V. Kryukov, F. Stegmeier, B.K. Erickson, L.
A. Garraway, W.R. Sellers, S.P. Gygi, Quantitative proteomics of the Cancer cell line encyclopedia, Cell 180 (2) (2020) 387–402, e16.
[55] C. Jean-Quartier, F. Jeanquartier, I. Jurisica, A. Holzinger, In silico cancer research towards 3R, BMC Cancer 18 (1) (2018) 408.
[56] S.T. Kim, R. Cristescu, A.J. Bass, K.-M. Kim, J.I. Odegaard, K. Kim, X.Q. Liu,
X. Sher, H. Jung, M. Lee, S. Lee, S.H. Park, J.O. Park, Y.S. Park, H.Y. Lim, H. Lee,
M. Choi, A. Talasaz, P.S. Kang, J. Cheng, A. Loboda, J. Lee, W.K. Kang, Comprehensive molecular characterization of clinical responses to PD-1 inhibition in metastatic gastric cancer %, J Nat. Med. 24 (9) (2018) 1449–1458.
[57] H. Dong, G. Zhu, K. Tamada, L. Chen, B7-H1, a third member of the B7 family, co- stimulates T-cell proliferation and interleukin-10 secretion, Nat. Med. 5 (12)
(1999) 1365–1369.
[58] H. Borghaei, L. Paz-Ares, L. Horn, D.R. Spigel, M. Steins, N.E. Ready, L.Q. Chow,
E.E. Vokes, E. Felip, E. Holgado, F. Barlesi, M. Kohlhaufl, O. Arrieta, M.A. Burgio, J. Fayette, H. Lena, E. Poddubskaya, D.E. Gerber, S.N. Gettinger, C.M. Rudin,
N. Rizvi, L. Crino, G.R. Blumenschein Jr., S.J. Antonia, C. Dorange, C.T. Harbison, F. Graf Finckenstein, J.R. Brahmer, Nivolumab versus docetaxel in advanced nonsquamous non-small-cell lung cancer, N. Engl. J. Med. 373 (17) (2015) 1627–1639.
[59] E.B. Garon, N.A. Rizvi, R. Hui, N. Leighl, A.S. Balmanoukian, J.P. Eder,
A. Patnaik, C. Aggarwal, M. Gubens, L. Horn, E. Carcereny, M.J. Ahn, E. Felip, J.
S. Lee, M.D. Hellmann, O. Hamid, J.W. Goldman, J.C. Soria, M. Dolled-Filhart, R.
Z. Rutledge, J. Zhang, J.K. Lunceford, R. Rangwala, G.M. Lubiniecki, C. Roach, K. Emancipator, L. Gandhi, K.-Investigators, Pembrolizumab for the treatment of non-small-cell lung cancer, N. Engl. J. Med. 372 (21) (2015) 2018–2028.
[60] P. Xu, H. Luo, Y. Kong, W.-F. Lai, L. Cui, X. Zhu, Cancer neoantigen: boosting immunotherapy, Biomed. Pharmacother. 131 (2020), 110640.
[61] S. Tan, D. Li, X. Zhu, Cancer immunotherapy: pros, cons and beyond, Biomed.
Pharmacother. 124 (2020), 109821.
[62] Z. Tang, D. Li, S. Hou, X. Zhu, The cancer exosomes: clinical implications, applications and challenges, Int. J. Cancer 146 (11) (2020) 2946–2959.
[63] C. Berthon, V. Driss, J. Liu, K. Kuranda, X. Leleu, N. Jouy, D. Hetuin, B. Quesnel, In acute myeloid leukemia, B7-H1 (PD-L1) protection of blasts from cytotoxic T cells is induced by TLR ligands and interferon-gamma and can be reversed using MEK inhibitors, Cancer Immunol. Immunother. 59 (12) (2010) 1839–1849.
[64] A.T. Parsa, J.S. Waldron, A. Panner, C.A. Crane, I.F. Parney, J.J. Barry, K.
E. Cachola, J.C. Murray, T. Tihan, M.C. Jensen, P.S. Mischel, D. Stokoe, R.
O. Pieper, Loss of tumor suppressor PTEN function increases B7-H1 expression and immunoresistance in glioma, Nat. Med. 13 (1) (2007) 84–88.
[65] K. Kataoka, Y. Shiraishi, Y. Takeda, S. Sakata, M. Matsumoto, S. Nagano, T. Maeda, Y. Nagata, A. Kitanaka, S. Mizuno, H. Tanaka, K. Chiba, S. Ito,
Y. Watatani, N. Kakiuchi, H. Suzuki, T. Yoshizato, K. Yoshida, M. Sanada,
H. Itonaga, Y. Imaizumi, Y. Totoki, W. Munakata, H. Nakamura, N. Hama,
K. Shide, Y. Kubuki, T. Hidaka, T. Kameda, K. Masuda, N. Minato, K. Kashiwase,

K. Izutsu, A. Takaori-Kondo, Y. Miyazaki, S. Takahashi, T. Shibata, H. Kawamoto,
Y. Akatsuka, K. Shimoda, K. Takeuchi, T. Seya, S. Miyano, S. Ogawa, Aberrant PD-L1 expression through 3'-UTR disruption in multiple cancers, Nature 534 (7607) (2016) 402–406.

[66] M.A. Coelho, S. de Carne Trecesson, S. Rana, D. Zecchin, C. Moore, M. Molina- Arcas, P. East, B. Spencer-Dene, E. Nye, K. Barnouin, A.P. Snijders, W.S. Lai, P. J. Blackshear, J. Downward, Oncogenic RAS signaling promotes tumor immunoresistance by stabilizing PD-L1 mRNA, Immunity 47 (6) (2017) 1083–1099, e6.

[67] C. Luchini, F. Bibeau, M.J.L. Ligtenberg, N. Singh, A. Nottegar, T. Bosse, R. Miller, N. Riaz, J.Y. Douillard, F. Andre, A. Scarpa, ESMO recommendations on microsatellite instability testing for immunotherapy in cancer, and its relationship with PD-1/PD-L1 expression and tumour mutational burden: a systematic review-based approach, Ann. Oncol. 30 (8) (2019) 1232–1243.

[68] R.S. Finn, M. Ikeda, A.X. Zhu, M.W. Sung, A.D. Baron, M. Kudo, T. Okusaka,
M. Kobayashi, H. Kumada, S. Kaneko, M. Pracht, K. Mamontov, T. Meyer,
T. Kubota, C.E. Dutcus, K. Saito, A.B. Siegel, L. Dubrovsky, K. Mody, J.M. Llovet, Phase ib study of lenvatinib plus pembrolizumab in patients with unresectable hepatocellular carcinoma, J. Clin. Oncol. 38 (26) (2020) 2960–2970.

[69] M.H. Taylor, C.H. Lee, V. Makker, D. Rasco, C.E. Dutcus, J. Wu, D.E. Stepan, R. C. Shumaker, R.J. Motzer, Phase IB/II trial of lenvatinib plus pembrolizumab in patients with advanced renal cell carcinoma, endometrial Cancer, and other selected advanced solid tumors, J. Clin. Oncol. 38 (11) (2020) 1154–1163. [70] R.K. Yuen, B. Thiruvahindrapuram, D. Merico, S. Walker, K. Tammimies, N. Hoang, C. Chrysler, T. Nalpathamkalam, G. Pellecchia, Y. Liu, M.J. Gazzellone,
L. D'Abate, E. Deneault, J.L. Howe, R.S. Liu, A. Thompson, M. Zarrei, M. Uddin, C.R. Marshall, R.H. Ring, L. Zwaigenbaum, P.N. Ray, R. Weksberg, M.T. Carter, B. A. Fernandez, W. Roberts, P. Szatmari, S.W. Scherer, Whole-genome sequencing of quartet families with autism spectrum disorder, Nat. Med. 21 (2) (2015) 185–191.

[71] K.L. Huang, E. Marcora, A.A. Pimenova, A.F. Di Narzo, M. Kapoor, S.C. Jin, O. Harari, S. Bertelsen, B.P. Fairfax, J. Czajkowski, et al., A common haplotype lowers PU.1 expression in myeloid cells and delays onset of Alzheimer's disease, Nat. Neurosci. 20 (8) (2017) 1052–1061.

[72] D. Gate, N. Saligrama, O. Leventhal, A.C. Yang, M.S. Unger, J. Middeldorp, K. Chen, B. Lehallier, D. Channappa, et al., Clonally expanded CD8 T cells patrol the cerebrospinal fluid in Alzheimer's disease, Nature 577 (7790) (2020) 399–404.

[73] P.R. Jansen, K. Watanabe, S. Stringer, N. Skene, J. Bryois, A.R. Hammerschlag, C. A. de Leeuw, J.S. Benjamins, A.B. Munoz-Manchado, M. Nagel, J.E. Savage, et al., Genome-wide analysis of insomnia in 1,331,010 individuals identifies new risk loci and functional pathways, Nat. Genet. 51 (3) (2019) 394–403.

[74] J.M. Lane, S.E. Jones, H.S. Dashti, A.R. Wood, K.G. Aragam, V.T. van Hees, L. B. Strand, B.S. Winsvold, H. Wang, J. Bowden, Y. Song, et al., Biological and clinical insights from genetics of insomnia symptoms, Nat. Genet. 51 (3) (2019) 387–393.

[75] D. Gill, M.K. Georgakis, F. Koskeridis, L. Jiang, Q. Feng, W.Q. Wei,
E. Theodoratou, P. Elliott, J.C. Denny, R. Malik, E. Evangelou, A. Dehghan,
M. Dichgans, I. Tzoulaki, Use of genetic variants related to antihypertensive drugs to inform on efficacy and side effects, Circulation 140 (4) (2019) 270–279. [76] O. Franzen, R. Ermel, A. Cohain, N.K. Akers, A. Di Narzo, H.A. Talukdar, H. Foroughi-Asl, C. Giambartolomei, J.F. Fullard, K. Sukhavasi, S. Koks, L.M. Gan,
C. Giannarelli, J.C. Kovacic, C. Betsholtz, B. Losic, T. Michoel, K. Hao, P. Roussos, J. Skogsberg, A. Ruusalepp, E.E. Schadt, J.L. Bjorkegren, Cardiometabolic risk loci share downstream cis- and trans-gene regulation across tissues and diseases, Science 353 (6301) (2016) 827–830.

[77] M. Bonaccio, A. Di Castelnuovo, S. Costanzo, E. Ruggiero, A. De Curtis,
M. Persichillo, C. Tabolacci, F. Facchiano, C. Cerletti, M.B. Donati, G. de Gaetano, L. Iacoviello, I. Moli-sani Study, Chili pepper consumption and mortality in italian adults, J. Am. Coll. Cardiol. 74 (25) (2019) 3139–3149.

[78] J.M. Havrilla, B.S. Pedersen, R.M. Layer, A.R. Quinlan, A map of constrained coding regions in the human genome, Nat. Genet. 51 (1) (2019) 88–95.

[79] S.M. Schussler-Fiorenza Rose, K. Contrepois, K.J. Moneghetti, W. Zhou, T. Mishra, S. Mataraso, O. Dagan-Rosenfeld, A.B. Ganz, J. Dunn, D. Hornburg, S. Rego, D. Perelman, S. Ahadi, M.R. Sailani, Y. Zhou, S.R. Leopold, J. Chen, M. Ashland, J.
W. Christle, M. Avina, P. Limcaoco, C. Ruiz, M. Tan, A.J. Butte, G.M. Weinstock, G.M. Slavich, E. Sodergren, T.L. McLaughlin, F. Haddad, M.P. Snyder, A longitudinal big data approach for precision health, Nat. Med. 25 (5) (2019) 792–804.

[80] N.S. Abul-Husn, K. Manickam, L.K. Jones, E.A. Wright, D.N. Hartzel, C. Gonzaga- Jauregui, C. O'Dushlaine, J.B. Leader, H. Lester Kirchner, D.M. Lindbuchler, M. L. Barr, M.A. Giovanni, M.D. Ritchie, J.D. Overton, J.G. Reid, R.P. Metpally, A.
H. Wardeh, I.B. Borecki, G.D. Yancopoulos, A. Baras, A.R. Shuldiner, O. Gottesman, D.H. Ledbetter, D.J. Carey, F.E. Dewey, M.F. Murray, Genetic identification of familial hypercholesterolemia within a single U.S. Health care system, Science 354 (6319) (2016).

[81] F.E. Dewey, M.F. Murray, J.D. Overton, L. Habegger, J.B. Leader, S.N. Fetterolf, C. O'Dushlaine, C.V. Van Hout, J. Staples, C. Gonzaga-Jauregui, et al., Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study, Science 354 (6319) (2016) aaf6814. [82] A.V. Khera, M. Chaffin, K.G. Aragam, M.E. Haas, C. Roselli, S.H. Choi, P. Natarajan, E.S. Lander, S.A. Lubitz, P.T. Ellinor, S. Kathiresan, Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations, Nat. Genet. 50 (9) (2018) 1219–1224.

[83] X. Yao, S. Tang, B. Bian, X. Wu, G. Chen, C.C. Wang, Improved phylogenetic resolution for Y-chromosome haplogroup O2a1c-002611, Sci. Rep. 7 (1) (2017) 1146.

[84] R. Ter Horst, M. Jaeger, S.P. Smeekens, M. Oosting, M.A. Swertz, Y. Li, V. Kumar, D.A. Diavatopoulos, A.F.M. Jansen, H. Lemmers, H. Toenhake-Dijkstra, A.E. van Herwaarden, M. Janssen, R.G. van der Molen, I. Joosten, F. Sweep, J.W. Smit, R.
T. Netea-Maier, M. Koenders, R.J. Xavier, J.W.M. van der Meer, C.A. Dinarello, N. Pavelka, C. Wijmenga, R.A. Notebaart, L.A.B. Joosten, M.G. Netea, Host and environmental factors influencing individual human cytokine responses, Cell 167 (4) (2016) 1111–1124, e13.

[85] B.J. Schmiedel, D. Singh, A. Madrigal, A.G. Valdovino-Gonzalez, B.M. White,
J. Zapardiel-Gonzalo, B. Ha, G. Altay, J.A. Greenbaum, G. McVicker, G. Seumois, A. Rao, M. Kronenberg, B. Peters, P. Vijayanand, Impact of genetic polymorphisms on human immune cell gene expression, Cell 175 (6) (2018) 1701–1715, e16.

[86] Q. Xiao, H. Yu, X. Zhu, The associations of hub gene polymorphisms in PI3K/ AKT/mTOR pathway and Schistosomiasis Japonica infection and hepatic fibrosis, Infect. Genet. Evol. 85 (2020), 104423.

[87] B. Liang, H. Ding, L. Huang, H. Luo, X. Zhu, GWAS in cancer: progress and challenges, Mol. Genet. Genomics 295 (3) (2020) 537–561.

[88] X. Zhu, J. Zhang, W. Fan, F. Wang, H. Yao, Z. Wang, S. Hou, Y. Tian, W. Fu,
D. Xie, W. Zhu, J. Long, L. Wu, X. Zheng, H. Kung, K. Zhou, M.C. Lin, H. Luo, D. Li, The rs391957 variant cis-regulating oncogene GRP78 expression contributes to the risk of hepatocellular carcinoma, Carcinogenesis 34 (6) (2013) 1273–1280.

[89] X. Zhu, M.C.M. Lin, W. Fan, L. Tian, J. Wang, S.S. Ng, M. Wang, H. Kung, D. Li, An intronic polymorphism in GRP78 improves chemotherapeutic prediction in non- small cell lung cancer, Chest 141 (6) (2012) 1466–1472.

[90] X. Zhu, M.S. Chen, L.W. Tian, D.P. Li, P.L. Xu, M.C. Lin, D. Xie, H.F. Kung, Single nucleotide polymorphism of rs430397 in the fifth intron of GRP78 gene and clinical relevance of primary hepatocellular carcinoma in Han Chinese: risk and prognosis, Int. J. Cancer 125 (6) (2009) 1352–1357.

[91] R. Vento-Tormo, M. Efremova, R.A. Botting, M.Y. Turco, M. Vento-Tormo, K.
B. Meyer, J.E. Park, E. Stephenson, K. Polanski, A. Goncalves, L. Gardner, S. Holmqvist, J. Henriksson, A. Zou, A.M. Sharkey, B. Millar, B. Innes, L. Wood,
A. Wilbrey-Clark, R.P. Payne, M.A. Ivarsson, S. Lisgo, A. Filby, D.H. Rowitch, J.

N. Bulmer, G.J. Wright, M.J.T. Stubbington, M. Haniffa, A. Moffett, S.
A. Teichmann, Single-cell reconstruction of the early maternal-fetal interface in humans, Nature 563 (7731) (2018) 347–353.
[92] C. Liu, N. Zhou, M.X. Du, Y.T. Sun, K. Wang, Y.J. Wang, D.H. Li, H.Y. Yu, Y. Song,
B.B. Bai, Y. Xin, L. Wu, C.Y. Jiang, J. Feng, H. Xiang, Y. Zhou, J. Ma, J. Wang, H. W. Liu, S.J. Liu, The mouse gut microbial biobank expands the coverage of cultured bacteria, Nat. Commun. 11 (1) (2020) 79.
[93] S. Greenblum, R. Carr, E. Borenstein, Extensive strain-level copy-number variation across human gut microbiome species, Cell 160 (4) (2015) 583–594.
[94] B.B. Finlay, C. Humans, Microbiome, are noncommunicable diseases communicable? Science 367 (6475) (2020) 250–251.
[95] B. Xie, D. Sun, Y. Du, J. Jia, S. Sun, J. Xu, Y. Liu, C. Xiang, S. Chen, H. Xie, et al., A two-step lineage reprogramming strategy to generate functionally competent human hepatocytes from fibroblasts, Cell Res. 29 (9) (2019) 696–710.
[96] D.K. Goode, N. Obier, M.S. Vijayabaskar, A.L.M. Lie, A.J. Lilly, R. Hannah,
M. Lichtinger, K. Batta, M. Florkowska, R. Patel, M. Challinor, K. Wallace,
J. Gilmour, S.A. Assi, P. Cauchy, M. Hoogenkamp, D.R. Westhead, G. Lacaud, V. Kouskoff, B. Gottgens, C. Bonifer, Dynamic gene regulatory networks drive hematopoietic specification and differentiation, Dev. Cell 36 (5) (2016) 572–587. [97] M.A. Ansari, V. Pedergnana, L.C.I.C.A. Magri, A. Von Delft, D. Bonsall, N. Chaturvedi, I. Bartha, D. Smith, G. Nicholson, G. McVean, A. Trebes, P. Piazza,
J. Fellay, G. Cooke, G.R. Foster, S.-H. Consortium, E. Hudson, J. McLauchlan, P. Simmonds, R. Bowden, P. Klenerman, E. Barnes, C.C.A. Spencer, Genome-to-genome analysis highlights the effect of the human innate and adaptive immune systems on the hepatitis C virus, Nat. Genet. 49 (5) (2017) 666–673.
[98] B. Liang, H. Yu, L. Huang, H. Luo, X. Zhu, A prognostic model guides surgical resection in cervical squamous cell carcinoma, Transl. Cancer Res. 9 (3) (2020) 1711–1731.
[99] J. Tang, Z. Xu, L. Huang, H. Luo, X. Zhu, Transcriptional regulation in model organisms: recent progress and clinical implications, Open Biol. 9 (11) (2019), 190183.
[100] D. Lu, Y. Huang, Y. Kong, T. Tao, X. Zhu, Gut microecology: why our microbes could be key to our health, Biomed. Pharmacother. 131 (2020), 110784. [101] M.A. Oberhardt, R. Zarecki, S. Gronow, E. Lang, H.P. Klenk, U. Gophna, E. Ruppin, Harnessing the landscape of microbial culture media to predict new organism-media pairings, Nat. Commun. 6 (2015) 8493.
[102] E.J. Topol, High-performance medicine: the convergence of human and artificial intelligence, Nat. Med. 25 (1) (2019) 44–56.
[103] G. Liang, W. Fan, H. Luo, X. Zhu, The emerging roles of artificial intelligence in cancer drug development and precision therapy, Biomed. Pharmacother. 128 (2020), 110255.
[104] T. Huang, L. Lan, X. Fang, P. An, J. Min, F. Wang, Promises and challenges of big data computing in health sciences, Big Data Res. 2 (1) (2015) 2–11.
[105] S. Wang, T. Huang, Applications of network analysis in biomedicine, Methods Mol. Biol. 2204 (2020) 39–50.
[106] W. Nelson, M. Zitnik, B. Wang, J. Leskovec, A. Goldenberg, R. Sharan, To embed or not: network embedding as a paradigm in computational biology, Front. Genet. 10 (2019) 381.
[107] Y. Gurovich, Y. Hanani, O. Bar, G. Nadav, N. Fleischer, D. Gelbman, L. Basel- Salmon, P.M. Krawitz, S.B. Kamphausen, M. Zenker, L.M. Bird, K.W. Gripp, Identifying facial phenotypes of genetic disorders using deep learning, Nat. Med. 25 (1) (2019) 60–64.
[108] S. Ravizza, T. Huschto, A. Adamov, L. Bohm, A. Busser, F.F. Flother, R. Hinzmann,
H. Konig, S.M. McAhren, D.H. Robertson, T. Schleyer, B. Schneidinger, W. Petrich, Predicting the early risk of chronic kidney disease in patients with diabetes using real-world data, Nat. Med. 25 (1) (2019) 57–59.
[109] A. Holzinger, P. Kieseberg, A.M. Tjoa, E. Weippl, Current Advances, Trends and Challenges of Machine Learning and Knowledge Extraction: from Machine Learning to Explainable AI, 2018, pp. 1–8, https://doi.org/10.1007/978-3-319- 99740-7 (Chapter 1).
[110] S. Wang, T.J.S.M. Huang, Applications of Deep Learning in Biomedicine, 1, 2021, pp. 29–39.
[111] A. Holzinger, G. Langs, H. Denk, K. Zatloukal, Mining H.M.J.W.I.R.D,
K. Discovery, Causability and explainability of artificial intelligence in medicine, Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 9 (4) (2019) e1312.
[112] A.J.B.I. Holzinge, Interactive machine learning for health informatics: when do we need the human-in-the-loop? Brain Inform. 3 (2) (2016) 119–131.
[113] J. Ranstam, M. Buyse, S.L. George, S. Evans, N.L. Geller, B. Scherrer, E. Lesaffre, G. Murray, L. Edler, J.L. Hutton, T. Colton, P. Lachenbruch, Fraud in medical research: an international survey of biostatisticians, ISCB Subcommittee Fraud, Control Clin Trials 21 (5) (2000) 415–427.
[114] I.J. Goodfellow, J. Shlens, C.S.J.C. Science, Explaining and Harnessing Adversarial Examples, 2014, pp. 1–10, arXiv e-Print archive.
[115] G. Massaro, C.N.Z. Mattar, A.M.S. Wong, E. Sirka, S.M.K. Buckley, B.R. Herbert, S. Karlsson, D.P. Perocheau, D. Burke, S. Heales, A. Richard-Londt, S. Brandner,
M. Huebecker, D.A. Priestman, F.M. Platt, K. Mills, A. Biswas, J.D. Cooper, J.K. Y. Chan, S.H. Cheng, S.N. Waddington, A.A. Rahim, Fetal gene therapy for neurodegenerative disease of infants, Nat. Med. 24 (9) (2018) 1317–1323.
[116] P.G. Holt, C.A. Jones, The development of the immune system during pregnancy and early life, Allergy 55 (8) (2000) 688–697.
[117] X. Tang, Y. Huang, J. Lei, H. Luo, X. Zhu, The single-cell sequencing: new developments and medical applications, Cell Biosci. 9 (2019) 53.
[118] B. Lin, L. Du, H. Li, X. Zhu, L. Cui, X. Li, Tumor-infiltrating lymphocytes: warriors fight against tumors powerfully, Biomed. Pharmacother. 132 (2020), 110873.
[119] A. Medvinsky, S. Rybtsov, S. Taoudi, Embryonic origin of the adult hematopoietic system: advances and questions, Development 138 (6) (2011) 1017–1031.
[120] D.M. Popescu, R.A. Botting, E. Stephenson, K. Green, S. Webb, L. Jardine, E. F. Calderbank, K. Polanski, I. Goh, M. Efremova, M. Acres, D. Maunder, P. Vegh,
Y. Gitton, J.E. Park, R. Vento-Tormo, Z. Miao, D. Dixon, R. Rowell, D. McDonald,
J. Fletcher, E. Poyner, G. Reynolds, M. Mather, C. Moldovan, L. Mamanova,
F. Greig, M.D. Young, K.B. Meyer, S. Lisgo, J. Bacardit, A. Fuller, B. Millar,
B. Innes, S. Lindsay, M.J.T. Stubbington, M.S. Kowalczyk, B. Li, O. Ashenberg,
M. Tabaka, D. Dionne, T.L. Tickle, M. Slyper, O. Rozenblatt-Rosen, A. Filby,
P. Carey, A.C. Villani, A. Roy, A. Regev, A. Chedotal, I. Roberts, B. Gottgens, S. Behjati, E. Laurenti, S.A. Teichmann, M. Haniffa, Decoding human fetal liver haematopoiesis, Nature 574 (7778) (2019) 365–371.
[121] Y. Erlich, D. Zielinski, DNA Fountain enables a robust and efficient storage architecture, Science 355 (6328) (2017) 950–954.
[122] C. Song, Y. Kong, L. Huang, H. Luo, X. Zhu, Big data-driven precision medicine: environment interactions inferred from genetic effects on phenotypic variability starting the custom-made era of iatrology, Biomed. Pharmacother. 129 (2020), in the UK Biobank, Sci. Adv. 5 (8) (2019) eaaw3538.
110445.
[125] S.E. Jones, J.M. Lane, A.R. Wood, V.T. van Hees, J. Tyrrell, R.N. Beaumont, A.
[123] Z. Zou, T. Tao, H. Li, X. Zhu, mTOR signaling pathway and mTOR inhibitors in R. Jeffries, H.S. Dashti, M. Hillsdon, K.S. Ruth, et al., Genome-wide association cancer: progress and challenges, Cell Biosci. 10 (2020) 31. analyses of chronotype in 697,828 individuals provides insights into circadian [124] H. Wang, F. Zhang, J. Zeng, Y. Wu, K.E. Kemper, A. Xue, M. Zhang, J.E. Powell, rhythms, Nat. Commun. 10 (1) (2019) 343.
M.E. Goddard, N.R. Wray, P.M. Visscher, A.F. McRae, J. Yang, Genotype-by-